

Online Decision Making for Trading Wind Energy

Miguel Angel Muñoz¹, Pierre Pinson^{2,3*} and Jalal Kazempour⁴

¹*OASYS Group, University of Malaga, Malaga, Spain.

²Dyson School of Design Engineering, Imperial College London, London, United Kingdom.

³Department of Technology, Management and Economics, Technical University of Denmark, Kgs. Lyngby, Denmark.

⁴Department of Wind and Energy Systems, Technical University of Denmark, Kgs. Lyngby, Denmark.

*Corresponding author(s). E-mail(s): p.pinson@imperial.ac.uk;
Contributing authors: miguelangeljmd@uma.es; jalal@dtu.dk;

Abstract

This paper proposes and develops a new algorithm for trading wind energy in electricity markets, within an online learning and optimization framework. In particular, we combine a component-wise adaptive variant of the gradient descent algorithm with recent advances in the feature-driven newsvendor model. This results in an online offering approach capable of leveraging data-rich environments, while adapting to nonstationary characteristics of energy generation and electricity markets, and with a minimal computational burden. The performance of our approach is analyzed based on several numerical experiments, showing both better adaptability to nonstationary uncertain parameters and significant economic gains.

Keywords: Decision making under uncertainty, Online learning, Electricity market, Newsvendor model

Nomenclature

Sets and Indices

j	Index of auxiliary information (features).
t	Index of time periods (hours).
\mathcal{T}^{in}	Training (in-sample) set.
\mathcal{T}^{oos}	Test (out-of-sample) set.

Parameters

ψ_t^+	Marginal opportunity cost for overproduction at hour t (€/MWh).
ψ_t^-	Marginal opportunity cost for underproduction at hour t (€/MWh).
\bar{E}	Maximum hourly wind energy production (MWh).
E_t	Actual wind energy produced at hour t (MWh).
\mathbf{x}_t	Vector of auxiliary information (features).
μ	Hyperparameter of the OLN algorithm.
η	Hyperparameter of the OLN algorithm that controls the learning rate.
α	Smoothing parameter of the OLN algorithm that controls the approximation to the original objective.
ρ	Decay constant of the OLN algorithm that controls the adaptation to new gradients.
ϵ	Constant strictly greater than zero.

Variables

\mathbf{q}_t	Decision rule vector of the OLN algorithm at hour t .
E_t^{F}	Energy bid for hour t of the market horizon submitted to the hour-ahead electricity market (MWh).
$\boldsymbol{\eta}_t$	Dynamic learning vector of the OLN algorithm at hour t .
$\eta_{t,j}$	Component j of $\boldsymbol{\eta}_t$ at hour t .
\mathbf{g}_t	Gradient or subgradient of the OLN objective function at hour t .
$g_{t,j}$	Component j of \mathbf{g}_t at hour t .

1 Introduction

1.1 Problem statement

Traditionally, the way in which trading wind energy has been considered relied on a two-step approach, starting with the predictive modeling of future energy generation (within either deterministic or probabilistic frameworks), and then the use of such forecasts as input to expected utility maximization strategies or, alternatively, some more general form of optimization problems, possibly also with stochastic frameworks and accommodating risk aversion. Although fruitful, these methodologies may be computationally expensive, while the value of

the final decisions is highly affected by the quality of the forecasts employed. It may then be beneficial to integrate the forecasting and decision-making stages, within a so-called *prescriptive analytics* framework (Bertsimas and Kallus, 2019). In parallel, electricity markets are amid rapid transformations towards reducing granularity and lead times, facilitating the integration of non-dispatchable energy sources but increasing the computational and adaptability requirements of the offering algorithms.

In a data-rich and nonstationary environment, approaches relying on online learning and online convex optimization are of direct relevance. For a very complete introduction to these topics, the reader is referred to Shalev-Shwartz et al (2012). On the one hand, online learning algorithms free the decision-maker from most assumptions about the wind or market dynamics. On the other hand, online learning algorithms are typically efficient methods capable of adapting to the increasing computational needs. Furthermore, the online learning analysis is based on regret as opposed to the classical maximization of the expected utility, possibly allowing to derive additional insights into the properties of trading strategies.

1.2 Status quo with trading wind energy and underlying newsvendor problems

Most wind energy is traded in wholesale electricity markets (referred to as forward markets in this paper), where an offer is submitted prior to the actual delivery of energy. However, the stochastic nature of wind energy entails incurring deviations from the original offer. There are countless ways of approaching this problem depending on the market structure and how uncertainty is addressed, and therefore it is infeasible to address such a vast literature. As a starting point, and since there is no single authoritative review that covers this topic of renewable energy offering in electricity markets, we refer the reader to Morales et al (2014) where the authors study different market variants and strategies assuming a classical stochastic programming framework, as well as Conejo et al (2010), which introduces general concepts of decision-making under uncertainty within electricity markets. We deal, in particular, with markets with a dual-price settlement for imbalances, under which there is no possibility of benefiting from a deviation and where imbalance penalties are asymmetric.

Early works in this area proposed an optimal quantile strategy based on probabilistic forecasts for wind energy production (Bremnes, 2004). Specifically, Pinson et al (2007) showed that, in its simplest version of a risk-neutral wind farms without any other assets (e.g., storage, conventional generation), the offering problem necessarily take the form of a newsvendor problem. Various generalizations were explored by others. Zugno et al (2013) proposed constraining the offer in both the power and probability space in order to accommodate risk aversion and behavioral aspects of trading (e.g., anchoring effect towards traditional single-valued forecasts). In parallel, Mazzi and Pinson (2016) devised and tested a reinforcement learning algorithm to track

the optimal quantile in a nonstationary environment. Similarly, [Dent et al \(2011\)](#) revisited the problem by accounting for the possibility of a population-based price-making behavior. And, for more complex versions of the offering problems, one can revert to a stochastic programming setup ([Morales et al, 2010](#)), for instance, owing to inter-temporal constraints, or risk-aversion. Even though these varied approaches explore alternative angles to generalizing the underlying newsvendor problems in wind energy offering in electricity markets, they still require a two-step procedure (predict, then optimize). In contrast, if employing a prescriptive approach such as the one described in the paper, since the forecasting step is overlooked, there is no need to provide a description of wind power generation and market quantities. There are hence no assumption involved in their dynamics.

Inspired by new advances in decision making under uncertainty in data-rich environments, this problem regained interest in recent years within a prescriptive analytics framework (hence, by integrating forecasting and optimization steps). [Stratigakos et al \(2022\)](#) used an ensemble of decision trees that considers the objective function to estimate the energy production. From the modeling perspective, the work of [Muñoz et al \(2020\)](#) is one of the closest to ours, also aligned with the new stream of research that utilizes features to produce context-specific decisions in a fully data-driven environment. Actually, they leveraged recent advances with data-driven newsvendor problems, recently covered by [Ban and Rudin \(2019\)](#), and proposed an approach that iteratively solves a linear optimization problem to update the generating offering decisions. Although relatively inexpensive, the time involved in its resolution may pose an issue in electricity markets like the Australian NEM¹, where trading and dispatching is based on 5-minute time steps and updates. Moreover, this approach seems redundant in the sense that the complete optimization problem is solved for each new step even though consecutive training sets may only differ by a few samples. These particular pitfalls pointed us to the thriving field of online learning in order to find suitable alternative approaches that are adaptive and computationally efficient.

1.3 Online learning vs. optimization

All the aforementioned references rely on direct optimization approach, using some forecast information as input. In contrast to such optimization approaches, learning is conceptually different. Indeed, in an optimization framework, based on a set of input, solving the optimization problem provides the optimal decisions directly. In an online learning framework, we have a succession of forecast-observations pairs that we use to iteratively update the parameters of a model. These parameters do not have to be the decision themselves, but instead the parameters of a model (or, policies) to generate proposal decisions as a function of input features.

The objective in an online learning paradigm is consequently not to minimize an objective function for a given instance of an optimization problem,

¹Australian National Electricity Market (NEM). See <https://aemo.com.au/>

but to minimize cumulative regret over the succession of decisions and related observations (with resulting consequences in the form of a loss). For an introduction to the subject, we refer the reader to the surveys of [Shalev-Shwartz et al \(2012\)](#) and [Hazan et al \(2016\)](#). Online Convex Optimization (OCO) focuses on the subset of online problems that face convex losses. Within OCO, we find a group of algorithms that continuously update the decision variables based on gradients or subgradients of a convex objective function, typically performing small but cumulative changes. These algorithms stand out for their versatility to deal with a wide range of problems and for their computational speed, provided that there is a closed-form expression to evaluate the (sub-)gradient ([Duchi et al, 2011](#); [Zheng, 2011](#)). These methods offer long-term regret guarantees ([Orabona, 2022](#)) and have proven to be helpful in numerous applications on power systems ([Gan and Low, 2016](#); [Hauswirth et al, 2017](#); [Colombino et al, 2019](#); [Guo et al, 2021](#); [Yuan et al, 2022](#)). The well-known online gradient descent, proposed by [Zinkevich \(2003\)](#), is the first in this category, achieving excellent results and giving rise to a wealth of derivative methods.

The strategy followed by the family of online gradient descent algorithms is in sharp contrast with most stochastic programming approaches, which solve an independent optimization problem with a different training set every time a decision has to be updated, e.g., the parameter of the decision rule in [Muñoz et al \(2020\)](#). Under convexity assumptions, an optimal solution can be found to each optimization problem, meaning that no single decision can ever achieve better performance on average in that training set. However, there is no certainty that the out-of-sample performance of such a decision enjoys the same privilege in finite sample sets. Instead, only probability guarantees can be offered even if the samples are i.i.d. ([Van Parys et al, 2021](#)).

Indeed, when samples are generated by a nonstationary environment following a process with seasonality or trend, the i.i.d. assumption does not hold and the out-of-sample performance can be very poor. This issue can be partly compensated using a rolling window setting ([Bashir and Lehtonen, 2018](#)) that updates the decision vector frequently. However, the dynamic process can evolve significantly within the training set. In that case, conventional stochastic programming approaches might still be hampered by unavoidably lengthy training set containing misleading old samples. On the contrary, online gradient descent algorithms update the decision through a point-wise update that involves the most recent information at the time, which enables capturing seasonality and trends even in short periods. Therefore, online gradient methods not only offer computational advantages but may also outperform their stochastic counterparts in terms of profits. In this line, the illustrative example provided in Section 4.2 and the case study presented in Section 5 corroborate these observations, showing faster tracking of nonstationary environments and an increase in the economic profit against rolling window approaches.

1.4 Contributions and structure

Far from being an isolated case, the Australian NEM is just an example of the existing trend pointing to a shortening in lead times and an increase in granularity, which is spreading across electricity markets. Without a doubt, these developments reduce the operational uncertainty, facilitating the integration of stochastic renewable energy sources², although, at the same time, they raise the computational needs and require methodologies that adapt to changes fast. To face these new challenges, we propose an algorithm that combines a feature-driven newsvendor model inspired by [Ban and Rudin \(2019\)](#) with a variant of the online gradient descent algorithm presented in [Zeiler \(2012\)](#). Intending to explore the full potential of the proposed algorithm and in line with this trend, we conceive a case study in which we analyze an hourly forward market that closes just before the start of the next period. This demonstration allows us to use actual data of the Danish Transmission System Operator (TSO), Energinet³, and provide a relevant test bench to demonstrate the salient features that online learning algorithms can bring to power system problems. To the best of our knowledge, this is the first paper that analyzes the problem of trading wind energy in the online learning setting. Next, we formulate the contribution of our work which are threefold:

- We develop an online offering algorithm for the wind farm problem merging powerful tools from online learning, decision making with contextual information and the technical literature of the wind farm. Results show that this algorithm is computationally inexpensive and achieves substantial economic profits.
- We propose a new nonstationary regret benchmark against which we empirically compare our algorithm.
- We showcase the ability of the proposed algorithm to adapt to nonstationary scenarios through a concise illustrative example, and demonstrate the superior economic performance and computational efficiency of our approach on a case study that leverages more than five years of realistic data published by the Danish TSO, Energinet.

The remaining of the manuscript is structured as follows: Section 2 introduces the problem of a wind farm offering in the forward market with a dual-price mechanism. Section 3 develops a new offering algorithm based on an adaptive gradient descent algorithm and explores several performance metrics. Section 4 is built upon two illustrative examples that investigate the behavior of an alternative online implementation and the dynamic response of this algorithm in comparison with previous rolling window approaches. Section 5 empirically analyzes the performance of our proposed algorithm in a case study that considers an hourly forward market based on real data retrieved from the Danish TSO, Energinet. Finally, conclusions are duly drawn in Section 6.

²Increasing time granularity in electricity markets, innovation landscape brief, International Renewable Energy Agency (IRENA), Report, 2019

³See <https://energinet.dk/>

2 Newsvendor problem on a rolling time-window

This section is devoted to introducing the problem of the wind farm offering in a forward market, which is cleared sometime before their actual production is realized. Therefore, the producer is likely to suffer deviations from her offer that are cleared *ex-post* in a real-time market under a dual-price financial settlement for imbalances. Furthermore, the offer is assumed to be always accepted, as the marginal operational cost of wind farms is close to zero and therefore this technology is usually prioritized for being scheduled. The market revenue $\rho \in \mathbb{R}$ of a wind farm in such a context is given by the summation of the amounts obtained in the forward (ρ^F) and in the real-time (also referred to as balancing) markets (ρ^B), i.e.,

$$\rho = \rho^F + \rho^B = \lambda^F E^F - \lambda^{\text{UP}}(E^F - E)^+ + \lambda^{\text{DW}}(E - E^F)^+, \quad (1)$$

where $(a)^+ = \max(a, 0)$. In addition, the unknown parameters $\lambda^F, \lambda^{\text{UP}}$, and $\lambda^{\text{DW}} \in \mathbb{R}$ are the forward, up-regulation and down-regulation prices, respectively. The key decision variable for the wind farm is her offer $E^F \in \mathbb{R}^+$ in the forward stage. Note that $E \in \mathbb{R}^+$ denotes the actual realization of her stochastic energy production, which is unknown in the forward stage. In accordance with (1), the revenue in the forward stage ($\lambda^F E^F$) is modified when the producer deviates from her offer E^F . When the production is greater than expected $E \geq E^F$, the producer needs to buy downward regulation at price λ^{DW} in the real-time stage to compensate the imbalance $E - E^F > 0$. On the contrary, if she produces less than her forward offer $E \leq E^F$, the wind farm has to buy upward regulation at price λ^{UP} to compensate the imbalance $E^F - E > 0$. Under a dual-price financial settlement, we have $\lambda^{\text{UP}} \geq \lambda^F$ and $\lambda^{\text{DW}} \leq \lambda^F$, with at most one of them different from λ^F (Morales et al, 2014, Ch. 7). In accordance with the aforementioned description, let $\psi^+, \psi^- \in \mathbb{R}^+$ denote penalties for over- or under-production as

$$\psi^+ = \lambda^F - \lambda^{\text{DW}}, \quad (2)$$

$$\psi^- = \lambda^{\text{UP}} - \lambda^F. \quad (3)$$

Using (2) and (3) and the equivalence $E - E^F = (E - E^F)^+ - (E^F - E)^+$, we reformulate (1) as

$$\rho = \lambda^F E - \left(\psi^+(E - E^F)^+ + \psi^-(E^F - E)^+ \right). \quad (4)$$

Note that the first term of (4) is out of the control of the underlying price-taker wind farm, as both λ^F and E are uncertain parameters. Therefore, the profit-maximizing offer E^{F*} of the wind farm in the forward market can be

computed by minimizing the expected deviation cost as

$$E^{\text{F}*} = \arg \min_{E^{\text{F}} \in [0, \bar{E}]} \mathbb{E} \left[\psi^+(E - E^{\text{F}})^+ + \psi^-(E^{\text{F}} - E)^+ \right], \quad (5)$$

where $\mathbb{E}[\cdot]$ is the expected operator. The optimization program (5) solves an instance of the very well-studied newsvendor model (Qin et al, 2011). Under a price-taker scenario, i.e., when the capacity of the producer is relatively small, an analytical solution to (5) can be computed through (Bremnes, 2004):

$$E^{\text{F}*} = F_E^{-1} \left(\frac{\bar{\psi}^+}{\bar{\psi}^+ + \bar{\psi}^-} \right), \quad (6)$$

where $F_E^{-1}(\cdot)$ is the cumulative distribution function (cdf) of the renewable energy production and the overline denotes the expected value of the random variable. The reader is referred to Maggioni et al (2019) for a discussion about the value of right distribution in newsvendor applications.

On top of the fact that the true distribution of the wind production and the optimal quotient are generally unknown, (6) suffers from another major drawback, which is its inability to directly leverage additional information that may be available, e.g., the wind energy forecasts of neighboring areas. In fact, it is usually the case that the wind farm is equipped with a vector of *auxiliary information*, also known as *features* $\mathbf{x} \subseteq \mathcal{X} \in \mathbb{R}^p$ where p denotes the dimension of the feature vector. This feature vector may help explain the behavior of the uncertain parameters in (5). As proposed by Ban and Rudin (2019), this information can be exploited in newsvendor instances assuming that the optimal offer follows a linear decision rule of the form $E^{\text{F}} : \mathcal{X} \rightarrow \mathbb{R}$, $E^{\text{F}} = \mathbf{x}^{\top} \mathbf{q}$ with $\mathbf{q} \in \mathbb{R}^p$ being a new decision vector that determines the linear model. Note that the dimension of the decision vector \mathbf{q} has also been chosen equal to p . This decision rule can easily reproduce an intercept setting a component of the feature vector \mathbf{x} equal to one. Then, considering that a set of historical samples of the form $\{(E_t, \psi_t^-, \psi_t^+, \mathbf{x}_t), \forall t \in \mathcal{T}^{\text{in}}\}$ is available, we compute the best decision \mathbf{q}^{LP} for this set through the following linear program:

$$\mathbf{q}^{\text{LP}*} = \arg \min_{\mathbf{q}} \frac{1}{|\mathcal{T}^{\text{in}}|} \sum_{t \in \mathcal{T}^{\text{in}}} \psi_t^+ (E_t - \mathbf{x}_t^{\top} \mathbf{q})^+ + \psi_t^- (\mathbf{x}_t^{\top} \mathbf{q} - E_t)^+ \quad (7a)$$

$$\text{s.t. } 0 \leq \mathbf{x}_t^{\top} \mathbf{q} \leq \bar{E}, \quad \forall t \in \mathcal{T}^{\text{in}}, \quad (7b)$$

where $|\cdot|$ denotes the cardinality of a set. Note that this model does not implicitly assume a price-taker scenario. In fact, correlations between penalties and wind features may be captured in systems with a high penetration of this technology. Although the linear structure of the mapping may seem restrictive, more complex relationships can be obtained by transforming the feature space,

e.g., using a Taylor approximation (Ban and Rudin, 2019). Next, defining the box projection

$$\pi(\mathbf{x}, \mathbf{q}) = \min \left(\max(0, \mathbf{x}^\top \mathbf{q}), \bar{E} \right), \quad (8)$$

the optimal offer for a new piece of information $\mathbf{x}_{t'}$ can be computed as $E_{t'}^F = \pi(\mathbf{x}_{t'}, \mathbf{q}^{\text{LP}})$. As discussed in Muñoz et al (2020), when new points are incorporated into the dataset \mathcal{T}^{in} , the problem (7) can be iteratively solved to update the value of \mathbf{q}^{LP} . In the remaining of the manuscript, we refer to this approach as LP (from Linear Programming).

3 Online learning in newsvendor problems

In the Online Convex Optimization (OCO) framework, a decision-maker faces an online learning problem where iterative decisions are to be made. The cost of each decision is determined by a convex loss function $f_t : \mathbb{R}^{d_z} \rightarrow \mathbb{R}$ unknown beforehand. After a decision $\mathbf{z}_t \in Z \subseteq \mathbb{R}^{d_z}$ is made, the decision-maker learns f_t and pays $f_t(\mathbf{z}_t)$. Within OCO the Online Gradient Descent (OGD) algorithm, introduced by Zinkevich (2003), has proven to be very effective and versatile (Gan and Low, 2016; Narayanaswamy et al, 2012; Hauswirth et al, 2016; Nonhoff and Müller, 2020; Wood et al, 2021). Starting from an initial value, the OGD performs iterative updates \mathbf{z}_t based on (sub-)gradients of f_t , denoted as \mathbf{g}_t from hereon. The magnitude of the step is controlled through a variable learning rate η_t . On each round, the updated vector is forced to lie within the feasible region Z through the Euclidean projection. In the OGD we rely on just the last point learned to obtain a gradient, thus resulting in a computationally inexpensive method, especially if the gradient and projection can be computed through a closed-form expression.

The selection of the learning rate is of paramount importance giving rise to countless variants of the gradient descent algorithm. The original proposal by Zinkevich (2003) presents two main alternatives, namely, a variable and a fixed learning rate. In a dynamic environment, the classical choice $\eta_t \in \mathbb{R}^+$, $\eta_t \propto t^{-1/2}$ where \propto denotes the proportional operator, is not suitable due to the fact that $\lim_{t \rightarrow \infty} \eta_t = 0$, reducing the ability to track changes as t increases. Alternatively, one could select a fixed value $\eta_t = \eta$ that keeps this capacity unaltered but may lose the fast convergence that the initial high values of η_t provide. Regardless of the selection, both choices are scale-dependent and treat each component of the gradient vector equally. To tackle this, McMahan and Streeter (2010) and Duchi et al (2011) propose to use a component-wise adaptive rate $\boldsymbol{\eta}_t \in \mathbb{R}^p$ and $\eta_{t,j} = \eta (\sum_{k=1}^t g_{k,j}^2)^{-1/2}$ where $g_{t,j}$ is a component of the gradient vector $\mathbf{g}_t = [g_{t,1}, \dots, g_{t,j}, \dots, g_{t,p}]^\top$. As in the case of $\eta_t \propto t^{-1/2}$, the previous expression is monotonically decreasing (component-wise), again limiting the long-term ability to learn. Aware of this limitation, Zeiler (2012)

suggests an exponentially decaying average of the squared gradients to modulate the learning rate based on the most recent information. We leverage this gradient descent variant to implement our algorithm in Section 3.1.

In the online learning community, the *de facto* metric to evaluate the performance of a series of decision vectors $\mathbf{z}_1, \dots, \mathbf{z}_T$ is the regret $\mathcal{R}_T \in \mathbb{R}$. The regret provides a versatile and, in a sense, normalized metric to compare an algorithm through different problems with the advantage that little assumption is made about the oracle that generates the decisions. Traditionally, the benchmark used to compute regret is the best single action in hindsight that can be obtained as the solution to an offline optimization problem under perfect information. However, in a dynamic environment, this benchmark can be beaten easily. In Section 3.3 we propose an alternative benchmark more suitable for the nonstationary context of the wind energy problem.

3.1 Online newsvendor

In this section, we particularize the gradient descent introduced in the previous paragraphs to the context of the wind farm offering in a forward market, incorporating elements of the rolling window problem presented in Section 2. We name the resulting algorithm OLN_V (from OnLine NewsVendor). Contrary to the rolling window approach, the OLN_V algorithm updates \mathbf{q} based on the information provided by the last realization. The objective function (7a) when the set \mathcal{T}^{in} reduces to one sample yields

$$NV_t(\mathbf{q}) = \psi_t^+ (E_t - \mathbf{x}_t^\top \mathbf{q})^+ + \psi_t^- (\mathbf{x}_t^\top \mathbf{q} - E_t)^+. \quad (9)$$

The OLN_V method requires computing a gradient of the objective function, for which we analyze two alternative procedures in the following paragraphs.

The first approach is inspired by the work of Zheng (2011) on the pinball loss, a particular case of the objective function found in newsvendor models. Since the pinball loss is not strictly differentiable, the authors propose an alternative smooth approximation to ensure that computing gradients is always possible. Note that in our case the objective function (9) is not differentiable at $E_t = \mathbf{x}_t^\top \mathbf{q}$. Therefore, we first propose to circumvent this issue extending the approach in Zheng (2011) to the more general expression (9) that considers arbitrary (positive) penalties as

$$NV_{t,\alpha}(\mathbf{q}) = \psi_t^+ (E_t - \mathbf{x}_t^\top \mathbf{q}) + \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{-(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha}), \quad (10)$$

where $\alpha > 0$ is a parameter that controls the approximation and where higher values of this parameter result in smoother functions. The function $NV_{t,\alpha}$ is convex in \mathbf{q} and upper bounds NV_t for any value of \mathbf{q} as proven in Propositions 1 and 2 in Appendix A, respectively. Then, we derive a closed-form

solution to obtain gradients of (10), yielding

$$\nabla NV_{t,\alpha}(\mathbf{q}) = \left(-\psi_t^+ + (\psi_t^+ + \psi_t^-) \frac{1}{1 + e^{(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha}} \right) \mathbf{x}_t. \quad (11)$$

The second approach deals directly with the objective function as formulated in (9). Even though the original objective is not strictly differentiable, a variant of the OLN algorithm is readily applicable to subdifferentiable functions, provided that a subgradient can be computed instead (Orabona, 2022). In this case, the mapping that returns a subdifferential of (9) is given by

$$\partial NV_t(\mathbf{q}) = \begin{cases} -\psi_t^+ \mathbf{x}_t, & E_t - \mathbf{x}_t^\top \mathbf{q} > 0, \\ \psi_t^- \mathbf{x}_t, & E_t - \mathbf{x}_t^\top \mathbf{q} < 0, \\ [-\psi_t^+ \mathbf{x}_t, \psi_t^- \mathbf{x}_t], & E_t - \mathbf{x}_t^\top \mathbf{q} = 0. \end{cases} \quad (12)$$

Note that, when $E_t - \mathbf{x}_t^\top \mathbf{q} = 0$, any value in the interval $[-\psi_t^+ \mathbf{x}_t, \psi_t^- \mathbf{x}_t]$ is a legitimate subgradient belonging to $\partial NV_t(\mathbf{q})$. For the sake of simplicity and reproducibility, the implementation of our algorithm returns zero whenever this condition is fulfilled.

Once a gradient as in (11) or a subgradient as in (12) has been computed, the key step of OLN is to update \mathbf{q}_t using a multidimensional learning rate $\eta_t \in \mathbb{R}^p$ through

$$\mathbf{q}_{t+1} = \Pi(\mathbf{q}_t - \eta_t \circ \mathbf{g}_t, \mathbf{x}_t), \quad (13)$$

where \circ denotes the element-wise product, $\mathbf{g}_t = \nabla NV_{t,\alpha}(\mathbf{q}_t)$ or $\mathbf{g}_t = \partial NV_t(\mathbf{q}_t)$ depending on the implementation of OLN, and Π is a projection operator defined as $\Pi : \mathbb{R}^p \times \mathcal{X} \rightarrow \mathbb{R}^p$. Precisely, Π maps its arguments into the solution of the following optimization problem:

$$\Pi(\mathbf{o}, \mathbf{x}) = \arg \min_{\mathbf{q} \in Q(\mathbf{x})} \frac{1}{2} \|\mathbf{q} - \mathbf{o}\|_2, \quad (\mathcal{P})$$

where \mathbf{o} represents a candidate to update the decision vector and is computed $\mathbf{o} = \mathbf{q}_t - \eta_t \circ \mathbf{g}_t$. The feasible set in (\mathcal{P}) is defined by the set-valued mapping $Q : \mathcal{X} \rightrightarrows \mathbb{R}^p$, $Q(\mathbf{x}) = \{\mathbf{q} : 0 \leq \mathbf{x}^\top \mathbf{q} \leq \bar{E}\}$. Note that, for any input \mathbf{x} the output of Q is a convex region bounded by two parallel hyperplanes. As the Euclidean norm is used, a unique solution is guaranteed to exist for any instance of (\mathcal{P}) . Generally, the Euclidean projection of a point into a convex set requires solving a convex optimization problem, however the definition of

Q allows us to find a closed-form expression, yielding

$$\Pi(\mathbf{o}, \mathbf{x}) = \begin{cases} \mathbf{o}, & 0 \leq \mathbf{x}^\top \mathbf{o} \leq \bar{E}, \\ \mathbf{o} + \frac{\bar{E} - \mathbf{x}^\top \mathbf{o}}{\|\mathbf{x}\|_2^2} \mathbf{x}, & \mathbf{x}^\top \mathbf{o} > \bar{E}, \\ \mathbf{o} + \frac{-\mathbf{x}^\top \mathbf{o}}{\|\mathbf{x}\|_2^2} \mathbf{x}, & \mathbf{x}^\top \mathbf{o} < 0. \end{cases} \quad (14)$$

This reduces the resolution of the optimization problem (\mathcal{P}) to evaluating the above expression. Even though the operator Π guarantees the feasibility of \mathbf{q}_t under the realization \mathbf{x}_{t-1} , we need to resort to (8) setting $E_t^F = \pi(\mathbf{x}_t, \mathbf{q}_t)$ to guarantee E_t^F remains feasible for any new arbitrary \mathbf{x}_t .

The last remaining aspect is to compute the vector $\boldsymbol{\eta}_t$ following the ideas in Zeiler (2012). Let $\mathbf{g}_t = [g_{t,1}, \dots, g_{t,j}, \dots, g_{t,p}]^\top$ be a gradient or subgradient vector computed through (11) and (12). Then, we can define the squared running average of each component as

$$\bar{g}_{t,j}^2 = \rho \bar{g}_{t-1,j}^2 + (1 - \rho) g_{t,j}^2, \quad (15)$$

where $\rho \in [0, 1)$ is a decay constant and $\bar{g}_{0,j}^2 = 0$. The auxiliary variable $\bar{g}_{t,j}^2$ is then used to compute the independent learning rate applied to the associated decision vector component following

$$\eta_{t,j} = \frac{\eta}{\sqrt{\bar{g}_{t,j}^2 + \epsilon}}, \quad (16)$$

where $\epsilon \in \mathbb{R}^+$ helps better condition the denominator and $\eta > 0$ is a constant. We use the update given by (15) and (16) in the proposed OLN algorithm with the values $\epsilon = 10^{-6}$ and $\rho = 0.95$, as originally suggested in Zeiler (2012). The benefits of this update is twofold. On the one hand, OLN adapts each learning rate component to the scale of the incumbent feature. On the other, OLN is capable to track the most recent dynamic between the uncertain vector $[E_t, \psi_t^+, \psi_t^-]$ and the feature vector \mathbf{x}_t .

Gathering all previous steps together, the complete OLN algorithm particularized for the feature-driven wind energy trading problem is compiled in Algorithm 1.

Despite the fact we have considered a single wind farm in the derivation, the proposed OLN algorithm is general enough to be exploited for an aggregation of wind farms, or in general, for a portfolio of diverse renewable energy sources with uncertain production, just by combining the capacity and generation of the assets. Equally, the potential spatial correlation among production of wind farms does not affect the feasible region of the newsvendor model, and therefore does not complicate the OLN algorithm. On the contrary, adding storage to the generation portfolio forces the model to include intertemporal constraints that dramatically reshape the feasible region, implying that the current decision will affect future outcomes. In this case, the decision-maker

Algorithm 1 Online Newsvendor (OLNV)

Require: Initial values $\mathbf{q}_1 \in \mathbb{R}^p$, $\eta > 0$, $\rho \in [0, 1)$, $\epsilon \in \mathbb{R}^+$

- 1: Initialize $\bar{g}_{0,j}^2 = 0$, $\forall j$
- 2: **for** $t = 1$ to T **do**
- 3: Output \mathbf{q}_t
- 4: Receive \mathbf{x}_t
- 5: Compute $E_t^F = \pi(\mathbf{x}_t, \mathbf{q}_t)$
- 6: Receive NV_t and pay $NV_t(E_t^F)$
- 7: Set $\mathbf{g}_t = \nabla NV_{t,\alpha}(\mathbf{q}_t)$ or $\mathbf{g}_t = \partial NV_t(\mathbf{q}_t)$
- 8: Accumulate $\bar{g}_{t,j}^2 = \rho \bar{g}_{t-1,j}^2 + (1 - \rho)g_{t,j}^2$, $\forall j$
- 9: Compute $\eta_{t,j} = \eta(\bar{g}_{t,j}^2 + \epsilon)^{-1/2}$, $\forall j$
- 10: Update $\mathbf{q}_{t+1} = \Pi(\mathbf{q}_t - \eta_t \circ \mathbf{g}_t, \mathbf{x}_t)$ solving (\mathcal{P})
- 11: **end for**

can resort to classical dynamic programming (Hargreaves and Hobbs, 2012) or more advanced learning algorithms such as budget-constrained online learning (Liakopoulos et al, 2019; Sherman and Koren, 2021) or reinforcement learning algorithms (Kuznetsova et al, 2013; Sutton and Barto, 2018).

Finally, even if a population effect may be present for renewables in electricity markets (i.e., even if price-taker individually, the sum of individual actions of these producers may impact market outcomes), several wind power producers can effectively use the OLNv algorithm to improve the profitability of their offer within the same region. The fact that each competing producer has different contextual information available and may process it in alternative ways mitigates possible increases in the volatility of their outcomes that could arise from correlated generation.

3.2 Regularization through average penalty anchoring

In a mature electricity market under a dual-price mechanism, it is common that $\psi_t^+ = \psi_t^- = 0$ throughout a significant number of hours, meaning that load and generation are close to equilibrium in the system. From (9), it is evident that, in this situation, the producer experiences no cost no matter the deviation from the actual production. Moreover, the gradients computed through (9) are zero and therefore the variable vector \mathbf{q}_t is not updated, wasting information about the relationship between E_t^F and \mathbf{x}_t . On a different front, when penalties are different from zero, they typically exhibit random behavior with sharp spikes representing highly imbalanced scenarios which, in turn, yields destabilizing updates of the vector \mathbf{q}_t . To tackle both issues, we propose performing the following convex transformation of the original penalties:

$$\psi_t^{+'} = \mu\psi_t^+ + (1 - \mu)\bar{\psi}^+, \quad (17)$$

$$\psi_t^{-'} = \mu\psi_t^- + (1 - \mu)\bar{\psi}^-, \quad (18)$$

where $0 \leq \mu \leq 1$ and $\bar{\psi}^+, \bar{\psi}^- \in \mathbb{R}^+$ are the historical average penalties. This convex transformation is inspired by the concept of constraining the optimal offer around the point forecast developed in [Zugno et al \(2013\)](#), but unlike them, we do not impose hard constraints on the decision vector \mathbf{q}_t . Instead, we smooth the objective function using as anchor the sample average optimal market quantile determined by the average market penalties $\bar{\psi}^+$ and $\bar{\psi}^-$. To do so, we consider a convex combination of the original objective function (7a) with an additional term that minimizes such a quantile,

$$\begin{aligned} NV_t^R = & \mu \psi_t^+ (E_t - \mathbf{x}_t^\top \mathbf{q})^+ + \mu \psi_t^- (\mathbf{x}_t^\top \mathbf{q} - E_t)^+ \\ & + (1 - \mu) \bar{\psi}^+ (E_t - \mathbf{x}_t^\top \mathbf{q})^+ + (1 - \mu) \bar{\psi}^- (\mathbf{x}_t^\top \mathbf{q} - E_t)^+. \end{aligned} \quad (19)$$

Then, by means of (17) and (18), the original objective structure is recovered, i.e.,

$$NV_t^R = \psi_t^{+'} (E_t - \mathbf{x}_t^\top \mathbf{q})^+ + \psi_t^{-'} (\mathbf{x}_t^\top \mathbf{q} - E_t)^+. \quad (20)$$

Therefore, by replacing ψ_t^+, ψ_t^- with $\psi_t^{+'}, \psi_t^{-'}$ in the original objective function, we regularize the learning procedure adding no extra computational cost. When the available samples are not sufficient to provide reliable estimators of the true $\bar{\psi}^+$ and $\bar{\psi}^-$, the producer can resort to assume a balanced market with penalties $\bar{\psi}^+ = \bar{\psi}^- = 1$. Thus, by the selection of $\mu < 1$, provided that $\bar{\psi}^+, \bar{\psi}^- > 0$, the algorithm leverages the information contained in samples with both penalties equal to zero, potentially accelerating the convergence and obtaining smoother updates through the gradient. The same reasoning applies to the smooth objective function.

3.3 Performance evaluation

In order to assess the economic performance of an algorithm over a set of testing samples $\{(E_t, \psi_t^-, \psi_t^+, \mathbf{x}_t), \forall t \in \mathcal{T}^{\text{OOS}}\}$, we can use the average deviation cost. To ease the notation, we set $T = |\mathcal{T}^{\text{OOS}}|$. Consider that we are equipped with a vector of offers E_1^F, \dots, E_T^F obtained from (7) and (8) or from Algorithm 1 after leaning one by one the samples belonging to \mathcal{T}^{OOS} . We compute the average deviation cost as

$$NV^{\text{OOS}} = \frac{1}{T} \sum_{t \in \mathcal{T}^{\text{OOS}}} \psi_t^- (E_t - E_t^F)^+ + \psi_t^+ (E_t^F - E_t)^+. \quad (21)$$

The value of this metric *per se* gives limited information about how a particular method is performing. A natural benchmark is the score obtained when a forecast of the wind energy production (in the sense of minimizing the root mean square error) is directly used as an offer in the market. We refer to this method as FO (from FOrecast). Let $NV_{\text{FO}}^{\text{OOS}}$ be the deviation cost incurred by

FO. Leveraging this quantity, we redefine the original metric in relative terms, i.e.,

$$NV^{\text{oos}}(\%) = \frac{NV_{FO}^{\text{oos}} - NV^{\text{oos}}}{NV_{FO}^{\text{oos}}} \cdot 100. \quad (22)$$

In this manner, the metric is expressed as a percentage of the possible improvement where a value of 100% means perfect performance with zero deviation cost.

On a different front, in online learning problems the customary performance measure is the regret. Traditionally, the regret compares a sequence of decision $\mathbf{q}_1, \dots, \mathbf{q}_T$ against the best single vector in hindsight $\mathbf{q}^{\mathcal{H}}$. The latter is computed *ex-post* solving a problem analogous to (7) once the whole collection of samples belonging to \mathcal{T}^{oos} is known. Let $Q^{\mathcal{H}}$ be the intersection of all feasible sets $Q(\mathbf{x}_t)$, more precisely $Q^{\mathcal{H}} : \mathcal{X} \rightrightarrows \mathbb{R}^p$, $Q^{\mathcal{H}} = \{\mathbf{q} : 0 \leq \mathbf{x}_t^\top \mathbf{q} \leq \bar{E}, t \in \mathcal{T}^{\text{oos}}\}$. The *static* regret renders

$$\mathcal{R}_T^s = \sum_{t \in \mathcal{T}^{\text{oos}}} NV_t(\mathbf{q}_t) - \min_{\mathbf{q} \in Q^{\mathcal{H}}} \sum_{t \in \mathcal{T}^{\text{oos}}} NV_t(\mathbf{q}). \quad (23)$$

Given the assumption of a nonstationary environment, outperforming a constant $\mathbf{q}^{\mathcal{H}}$ can be a relatively easy task even though it is determined under perfect information. On the other side of the spectrum, one may consider the *worst-case* regret (Besbes et al, 2015) interchanging the sum and minimum, i.e.,

$$\mathcal{R}_T^w = \sum_{t \in \mathcal{T}^{\text{oos}}} NV_t(\mathbf{q}_t) - \sum_{t \in \mathcal{T}^{\text{oos}}} \min_{\mathbf{q} \in Q(\mathbf{x}_t)} NV_t(\mathbf{q}), \quad (24)$$

where the second term of (24) computes the best individual decision $\mathbf{q}_t^{\mathcal{H}} \in \arg \min_{\mathbf{q} \in Q(\mathbf{x}_t)} NV_t(\mathbf{q})$. The regret computed in this way can be very pessimistic and unrealistic. Note that in the context of the wind farm, it is always possible to find a value for \mathbf{q} such that $E_t - \mathbf{x}_t^\top \mathbf{q} = 0$, and therefore (24) readily reduces to the summation of the original objective function $\mathcal{R}_T^w = \sum_{t \in \mathcal{T}^{\text{oos}}} NV_t(\mathbf{q}_t)$. Alternatively, Zinkevich (2003) proposed to compare the performance of online algorithms against a sequence of arbitrary decisions $\mathbf{u}_1, \dots, \mathbf{u}_T$, $\mathbf{u}_t \in Q(\mathbf{x}_t)$,

$$\mathcal{R}_T^d = \sum_{t \in \mathcal{T}^{\text{oos}}} NV_t(\mathbf{q}_t) - \sum_{t \in \mathcal{T}^{\text{oos}}} NV_t(\mathbf{u}_t). \quad (25)$$

We refer to this approach as *dynamic* regret. This formulation allows to define a metric with an adjustable difficulty between the previous benchmarks. Note that (23) and (24) are special cases of (25) with $\mathbf{u}_t = \mathbf{q}^{\mathcal{H}} \forall t$ and $\mathbf{u}_t = \mathbf{q}_t^{\mathcal{H}} \forall t$, respectively. Then, the question is how to choose a reasonable series of comparators \mathbf{u}_t to use against OLN. To this end, we

propose dividing \mathcal{T}^{OOS} in k adjacent partitions of equal length l , except possibly the last one. Without loss of generality, by assuming $T - kl = 0$, we have $\mathcal{T}_i^{\text{OOS}} = \{t : (i - 1)l + 1 \leq t \leq il\}, i = 1, \dots, k$. Let us define the feasible sets $Q_i^{\mathcal{H}} = \{\mathbf{q} : 0 \leq \mathbf{x}_t^\top \mathbf{q} \leq \bar{E}, t \in \mathcal{T}_i^{\text{OOS}}\}$. Accordingly, we can compute $\mathbf{q}_i^{\mathcal{H}} = \arg \min_{\mathbf{q} \in Q_i^{\mathcal{H}}} \sum_{t \in \mathcal{T}_i^{\text{OOS}}} NV_t(\mathbf{q})$. Finally, the sequence of comparators that we propose to use in this paper is $\mathbf{u}_t = \mathbf{q}_i^{\mathcal{H}}, \forall t \in \mathcal{T}_i^{\text{OOS}}$. We will empirically investigate the regret performance of OLVN in the case study presented in Section 5.

4 Illustrative examples

This section analyzes several illustrative examples to gain insight into the behavior of OLVN. The first case compares the two alternative implementations introduced in Section 3.1 and discusses their main properties. As a result of this analysis, we select the subgradient objective function as the default procedure to perform the update of \mathbf{q}_t in OLVN. On a different front, one of the key features of online learning algorithms is their tracking ability, given the chronological order in which the updates are performed. In the second illustrative example, we deal with alternating penalty scenarios, demonstrating the salient properties of OLVN to adapt to a changing environment in a concise case.

4.1 Comparing the smooth and subgradient implementations

This illustrative example aims to elucidate whether the smooth approximation presented in (10) provides any advantage over the direct subgradient implementation of OLVN and determine the implementation to be used in the rest of the numerical experiments.

To this end, we consider a simplified setting with a single feature, a forecast of the actual wind production that we also use as the baseline of the method FO, and a single regressor variable $q_t \in \mathbb{R}$. No intercept is considered to ease the representation and analysis of q_t . We sample the feature from a uniform distribution $x_t \sim U(10, 90)$ (MW) and the true wind generation series is built adding a normal noise $E_t = x_t + \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, 6)$ (MW). We produce a dataset of 1 year (8760 samples) through this process. Given that the penalties ψ_t^+ and ψ_t^- are difficult to simulate, we compute them based on real day-ahead and regulation prices of the Danish DK1 bidding zone. We retrieve data corresponding to the year 2017 from the data portal of the Danish TSO, Energinet⁴. Four implementations of Algorithm 1 are executed, three of them computing gradients of the smooth objective function through (11) with $\alpha = 0.05, 5$ and 20 and the last one leveraging subgradients of the original cost mapping as in (12), to which we refer to as ∂ . All instances are initialized with $q_1 = 1$, which means that the first offer produced by FO and OLVN are the

⁴See <https://www.energidataservice.dk/>

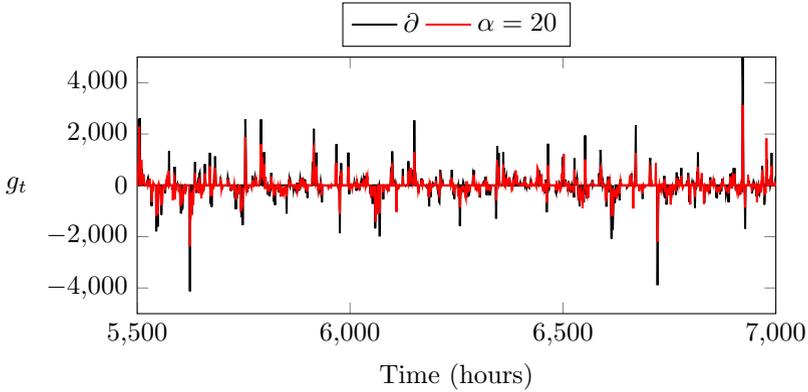


Fig. 1: Sample of ∂NV_t and $\nabla NV_{t,20}$ computed in the dataset of the illustrative example.

same. In this section we do not use any convex transformation of the prices, i.e., $\mu = 1$, and we set $\eta = 0.005$. We run the OLN algorithm throughout the dataset, performing updates of q_t every hour.

In this section, we accompany the numerical results with some theoretical analysis. The function $NV_{t,\alpha}(\mathbf{q})$ approximates well the original function $NV_t(\mathbf{q})$ when $|E_t - \mathbf{x}_t^\top \mathbf{q}| \rightarrow \infty$ as shown in Proposition 3 in Appendix A. Then, the question that remains is the behavior of both functions in a neighborhood of $E_t - \mathbf{x}_t^\top \mathbf{q} = 0$, defined $\varphi = \{\mathbf{q} : -\delta \leq E_t - \mathbf{x}_t^\top \mathbf{q} \leq \delta\}$ with $\delta > 0$ and δ close to zero. Let \mathbf{q}_1 and \mathbf{q}_2 be two vectors with $E_t - \mathbf{x}_t^\top \mathbf{q}_1 \leq 0$, $E_t - \mathbf{x}_t^\top \mathbf{q}_2 \geq 0$ and $\mathbf{q}_1, \mathbf{q}_2 \in \varphi$. The subgradient that OLN computes for each vector changes dramatically with $\partial NV_t(\mathbf{q}_1) = \psi_t^- \mathbf{x}_t$ and $NV_t(\mathbf{q}_2) = -\psi_t^+ \mathbf{x}_t$, which may result in diametrically opposed updates of the vector \mathbf{q} for similar values of \mathbf{x}_t or \mathbf{q}_t . Conversely, $NV_{t,\alpha}$ is everywhere differentiable, which ensures a smooth change of $\nabla NV_{t,\alpha}(\mathbf{q})$ for similar values of $\mathbf{q} - T$ and \mathbf{x}_t .

Figure 1 shows a sample of ∂NV_t and $\nabla NV_{t,20}$ that correspond to the subgradient and gradient of the smooth objective function with $\alpha = 20$, respectively. Note that only NV_t and $NV_{t,20}$ are represented for the sake of clarity. It can be observed that most of the spikes in the case of $NV_{t,20}$ are comparatively lower due to the aforementioned smoothing effect in the neighborhood of $E_t - \mathbf{x}_t^\top \mathbf{q} = 0$. This observation is aligned with the decreasing value of the standard deviation of the (sub-)gradients σ collated in Table 1 as α increases.

Table 1: Average absolute value $|\bar{g}|$ and standard deviation σ of the (sub-)gradients and the metric $NV^{\text{os}}(\%)$ computed for three smooth (α) and one subgradient (∂) implementations of the OLN.

	∂	$\alpha = 0.05$	$\alpha = 5$	$\alpha = 20$
$ \bar{g} $	121.7	122.0	125.6	133.5
σ	380.8	379.7	310.4	293.3
$NV^{\text{os}}(\%)$	5.3	5.2	0.8	-14.5

On the contrary, the mean absolute value of the (sub-)gradients, denoted as $|\bar{g}|$, follows the opposite evolution. To understand the rationale behind this evolution, we provide Figure 2 showing three instances of the original and smooth losses. In all cases, we corroborate that $NV_{t,\alpha}$ upper bounds NV_t by a finite amount as proved in Proposition 2 in Appendix A. However, plot 2(a) evidences that the minimum of $NV_{t,\alpha}$ is biased. This is true whenever $\psi_t^+ \neq \psi_t^-$, a common situation in markets with a dual-price settlement for imbalances. Furthermore, when one penalty is equal to zero, the minimum is never attained.

In this line, the gradient computed through (11) almost always introduces an error strictly greater than zero compared to the true value returned by (12). The value of this error is given by the following expression:

$$\begin{aligned} \nabla NV_{t,\alpha} - \partial NV_t(\mathbf{q}) = & \\ & \begin{cases} (\psi_t^+ + \psi_t^-)(1 + e^{(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha})^{-1} \mathbf{x}_t, & E_t - \mathbf{x}_t^\top \mathbf{q} > 0, \\ -(\psi_t^+ + \psi_t^-)(1 + e^{-(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha})^{-1} \mathbf{x}_t, & E_t - \mathbf{x}_t^\top \mathbf{q} < 0, \\ [-\frac{\psi_t^+ + \psi_t^-}{2} \mathbf{x}_t, \frac{\psi_t^+ + \psi_t^-}{2} \mathbf{x}_t], & E_t - \mathbf{x}_t^\top \mathbf{q} = 0. \end{cases} \end{aligned} \quad (26)$$

The imperfect approximation of $NV_{t,\alpha}$ distorts the magnitude and even the sign of the gradients, causing a long-term drift of q_t that increases with the smoothing parameter α as shown in Figure 3.

Finally, the last row of Table 1 presents the $NV^{\text{OOS}}(\%)$ obtained by each implementation with respect to FO. In this table, it is clear that $NV^{\text{OOS}}(\%)$ deteriorates when α increases. In summary, the smooth approach increasingly dampens the evolution of the decision vector for higher values of α but at the expense of a biased q_t and important economic losses. Therefore, we conclude that the smooth approximation does not provide any substantial advantage over the subgradient implementation in this application, given that the producer is neutral to risk and volatility (only concerned with the economic return), and there is no technical constraint that encourages a smooth evolution of q . As a result of this analysis, we use subgradients to implement the OLVN method throughout the rest of the paper.

4.2 Dynamic behavior

In this illustrative example, we compare the tracking ability of OLVN and LP approaches in a nonstationary environment. Similar to the previous case, we assume that the producer has access to a unique feature and considers a model with a single regressor. Again, we sample the forecast from a uniform distribution $x_t \sim U(10, 90)$ (MW) and the true wind generation series is built adding a normal noise $E_t = x_t + \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, 6)$ (MW). Instead of the real DK1 data, we consider two possible scenarios with penalties $\psi_t^+ = 1, \psi_t^- = 3$ and $\psi_t^+ = 3, \psi_t^- = 1$, alternating every two months. This process generates 8 months of data (5760 hours) using the last 4 months (2880 hours) as the test set. The start of the test set is aligned with the beginning of a two-month scenario with $\psi_t^+ = 1$ and $\psi_t^- = 3$. The rolling window approach

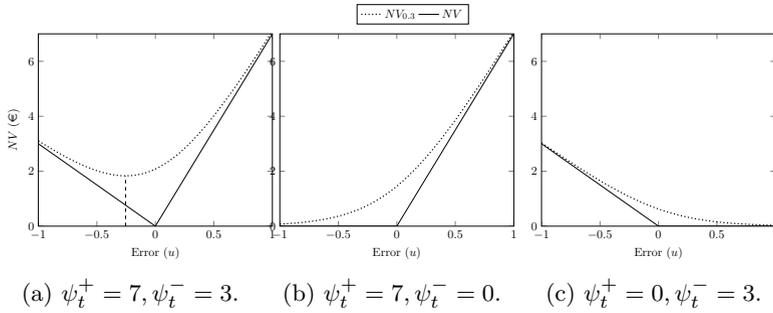


Fig. 2: Different instances of the original NV and smooth $NV_{0.3}$ objective function with $\alpha = 0.3$ and $u = E_t - x_t q$.

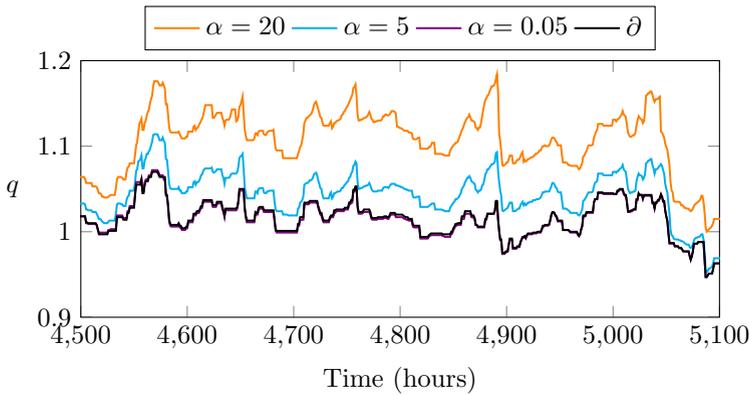


Fig. 3: Example of the evolution of the coefficient q for different implementations of OLNv.

is implemented solving the optimization problem (7) with a set of historical samples \mathcal{T}^{in} . Then, we leverage (8) to cast an offer based on the context $E_t^{\text{F}} = \pi(x_t, q_t^{\text{LP}})$. The coefficient q_t^{LP} is refreshed every 24 hours solving problem (7), following a rolling window. The reason for a 24-hour update is twofold: it is aligned with the original proposal in Muñoz et al (2020) and we empirically checked that there was little economic gain to be had with a lower update. The resolution time in the case of an hourly update, for example, took 24 times longer. As we later discuss, LP follows a rolling window approach that produces small changes in the training set, resulting in similar q_t^{LP} . We train four versions of the LP model with $|\mathcal{T}^{\text{in}}| = 720, 1440, 2160$ and 2880 (1, 2, 3, or 4 months), denoted as LP-1M to LP-4M, respectively. We use the first four months of the dataset to construct the initial training sets. Although the concept of training is substantially vaguer in OLNv, the last month of the training set is used to update the value of q_t , originally initialized as $q_1 = 1$, to resemble a model that has been operating for some time.

Figure 4 depicts the evolution of the single regressor q_t in the test set, together with the optimal q^* in each penalty scenario. In the first two months, the higher value of ψ_t^- penalizes offers above the true production $E_t^F > E_t$ and, consequently, the optimal strategy is to underestimate E_t^F with $q^* < 1$. In the final months, the opposite is the case.

As one may expect, the evolution of the decision vector of LP models is smoother than in the case of OLNv, given that the first one considers many historical samples to perform the update. However, Figure 4 shows that the trajectory of q_t produced by the rolling window models LP-1M to LP-4M is substantially lagged with respect to the change in the penalty scenario (emphasized by different background colors). This delay increases with the length of the training set to the point that LP-4M completely overlooks it. Note that the length of the training set in LP-4M and the period of the penalty scenarios are identical. Therefore, the number of samples that penalize under- or overproduction is equal and remains constant. As a result, LP-4M offers no incentive to overestimate or underestimate the forecast, yielding the same value as FO (neglecting slight deviations due to the finite sample and noise).

Figure 4 also shows that OLNv is substantially faster tracking the optimal q^* given that its updates are free from misleading past samples. In effect, the LP problem (7) determines the decision q_t with the best performance on average in the training set, assuming that all the samples in the set are equally probable representations of future outcomes. Conversely, OLNv only uses the most recent information to perform a point-wise update that swiftly captures changes in the environment. The tracking capability of each model has an impact on its economic performance. Table 2 summarizes the out-of-sample NV^{os} (%) obtained by each approach in the test set. In line with the previous analysis, LP-4M obtains the same performance as FO. The other three LP methods experience decreasing NV^{os} (%) as the length of the training set and the lag of q_t increase. Finally, the adaptability of OLNv clearly outperforms the LP approaches.

Table 2: Out-of-sample NV^{os} (%) obtained in the test set of the illustrative example.

	OLNV	LP-1M	LP-2M	LP-3M	LP-4M
NV^{os} (%)	13	5	-5	-6	0

In this simplified example, we could have analyzed LP models with a shorter training set, probably resulting in reduced lag and better performances. However, in a realistic situation with a huge feature space and random penalties, months of data are typically required to capture the underlying relationships and generalize well in the out-of-sample set (Muñoz et al, 2020). Therefore, the length of the training set of the LP models has to be selected as a trade-off; enough data is required to learn a policy that generalizes well, but shorter sets capture dynamics better. On the contrary, the OLNv approach completely

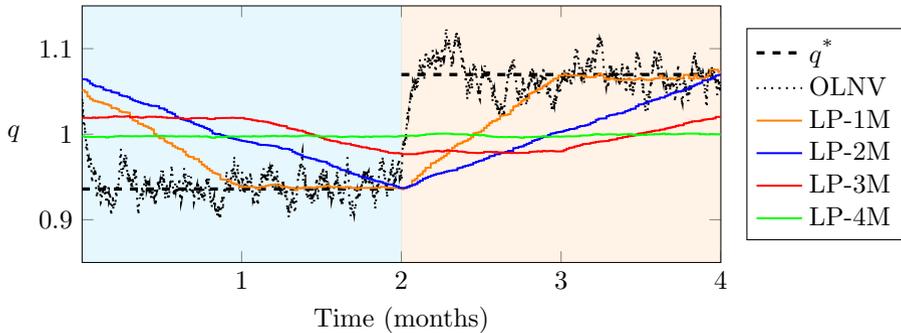


Fig. 4: Evolution of q produced by five models over the test set. The blue and orange shaded periods correspond to the penalty scenarios $\psi_t^+ = 1, \psi_t^- = 3$ and $\psi_t^+ = 3, \psi_t^- = 1$, respectively. The entry q^* corresponds to the best single vector for each penalty scenario.

avoids this dichotomy, providing a fast and effective method that adapts to uncertain parameters generated by nonstationary environment.

5 Case study

Electricity markets are in the midst of a rapid development towards reducing the time between market transactions and the actual exchange of electricity. Examples of this transformation are given, i.e., by the reduction of the electricity lead time (Australian NEM or the Californian CAISO⁵) or by the development of new intraday markets (OMIE intraday markets or NordPool ELBAS⁶). Inspired by this trend, we analyze a case study that considers an online forward market that takes place every hour followed by a balancing market with a dual-price settlement for imbalances. The closure of the forward market happens just before the start of the next period. We assume that the wind farm constantly participates in the market and her offer is always accepted.

This section first describes the data utilized in this case study. Then, several benchmark methods are proposed to compare against OLVN. Finally, in the last part of this section, we analyze the numerical results obtained in the case study, including the regret, economic performance and computational cost of OLVN.

5.1 Data and experimental setup

This case study uses historical data compiled by the Danish TSO, Energinet.dk, since it includes market prices and several wind power forecasts that can be leveraged as quality features. We collect the true and day-ahead

⁵See <https://aemo.com.au> and <http://www.caiso.com/>

⁶See <https://www.omie.es/> and <https://www.nordpoolgroup.com/>

forecast issued by Energinet for the on- and offshore wind power production of both DK1 and DK2 Danish bidding zones together with the day-ahead and regulation prices of DK1 for the period 01/07/2015 to 06/04/2021 (mm/dd/yyyy). The day-ahead spot and regulation prices are mapped into hourly penalties through equations (2) and (3) and some small negative values, obtained due to rounding errors, are filtered out.

Table 3: Installed capacity in MW by bidding zone and technology.

year	DK1		DK2	
	Onshore	Offshore	Onshore	Offshore
2015	2966	843	608	428
2016	2966	843	608	428
2017	2966	843	608	428
2018	3664	1277	759	423
2019	3669	1277	757	423
2020	3645	1277	757	423
2021	3725	1277	756	423

The raw wind power forecast series are also processed to be used in our case study. Given that the installed capacity of the four wind categories shown in Table 3 evolves differently over the dataset, we independently normalize each series to lie between 0 and 100 MW, a figure that can easily represent the capacity of a standard power plant. According to the Danish TSO, the raw wind power forecasts are issued between 12 to 36 hours ahead, although the exact time is difficult to know because no timestamp is provided. To overcome this issue, we use a standard ordinary least square regression model to produce enhanced forecasts with an accuracy comparable to an hour-ahead forecast and, therefore, suitable for our case study. We feed each raw wind power forecast into an independent linear regression model together with the last three lags of the true historical wind realization of the pertaining series. Finally, we use the first 6 months of our dataset to independently train each of the four predictive models, one per column of Table 3.

Table 4: Average RMSE (MWh) of the original forecast, the persistent (naive 1h lag) and improved 1h-ahead forecast computed on the out-of-sample period 07/01/2015 to 06/04/2021 with a normalized generation capacity of 100 MW.

Model	DK1		DK2	
	Onshore	Offshore	Onshore	Offshore
original	6.19	9.55	6.77	10.68
persistent	3.36	6.39	3.90	7.49
improved	2.72	5.70	3.34	6.66

Table 4 compares the root mean square error (RMSE) of the original and improved out-of-sample forecast against the naive benchmark provided by the first lag of each series (the wind power production of the previous hour), also

known in the literature as persistent estimator. Results show that the improved hour-ahead series significantly outperforms the original and persistent estimator. Hence, these enhanced forecasts are suitable for our case study. As a byproduct of this table, it is interesting to note that the wind forecasts issued by the Danish TSO have coherent RMSE being offshore harder to predict than onshore and DK2 harder than DK1 due to a reduced installed capacity and coverage area.

Once we have processed the wind power production series, we explain how we use them in our case study. The stochastic generation of the wind farm offering in the market is simulated using the normalized onshore wind data series of the Danish DK1 bidding zone, which is consistent with the bidding zone of the imbalance penalties utilized. The four hour-ahead forecasts of the wind power production of DK1 and DK2 are available to the producer as contextual information. Although additional wind forecasts of neighboring bidding zones could have been used as features, we restrict ourselves to the ones produced by the Danish TSO to avoid potential inconsistencies regarding the issuing time that could cast doubt on the results obtained (Muñoz et al, 2020).

Given that our goal is to reduce the imbalance cost incurred by the wind farm, we also consider several price-related features to be used as contextual information. To this end, we include the first lag of the imbalance penalties ψ_{t-1}^+ and ψ_{t-1}^- in the vector of contextual information. As commented in Section 2, it is well known that the ratio between the penalties provides valuable information about the optimal decision of the newsvendor model and, therefore, we add the series $r_{t-1} = \psi_{t-1}^+ / (\psi_{t-1}^+ + \psi_{t-1}^- + v)$ where $v = 10^{-5}$ is a constant that helps better condition the denominator. Finally, we add a column of ones that enable one of the regressors to become an intercept, completing our feature set.

As a summary, let $E_t^{on1}, E_t^{of1}, E_t^{on2}, E_t^{of2}$ denote the hour-ahead wind power forecast of DK1 onshore, DK1 offshore, DK2 onshore and DK2 offshore, respectively. Then, at the moment of delivering the offer, the producer has available a feature vector $\mathbf{x}_t = [1, E_t^{on1}, E_t^{of1}, E_t^{on2}, E_t^{of2}, \psi_{t-1}^+, \psi_{t-1}^-, r_{t-1}]^T$ to infer the optimal offer E_t^F .

5.2 Benchmark methods and implementation details

In this section, we describe several benchmark methods against which we compare the performance of OLN. The first opponent to beat is the enhanced hourly forecast of DK1 itself, produced through the ordinary least square regression model described before. Although a prediction that minimizes the RMSE may seem naive, one can expect that the deviation cost incurred by the producer vanishes as the RMSE of the forecast approaches zero. Therefore, an hour-ahead forecast is expected to perform relatively well. We also use this hour-ahead forecast as the baseline to compute the metric $NV^{oos}(\%)$ for the rest of the approaches in the way described in Section 3.3.

The second method discussed has been recently introduced by Muñoz et al (2020). In this work, the authors propose a two-step approach leveraging two

variants of model (7). In the first step, the first model only considers wind-related features plus the intercept and set $\psi_t^+ = \psi_t^- = 1, \forall t$. The series resulting from such model can be interpreted as an enhanced forecast of the wind energy production with a reduced mean absolute error. In a second step, this enhanced forecast is fed into model (7), considering this time the true historical penalties ψ_t^+ and ψ_t^- to correct for market patterns but neglecting the capacity constraint (7b). The training set is updated following a rolling window, adding new samples and eliminating the same amount of the oldest. We replicate this method, called LP2 (Linear Programming 2-steps), considering the four hour-ahead enhanced wind power forecasts of DK1 and DK2 as the input of the first step, this is, $\mathbf{x}_t = [1, E_t^{on1}, E_t^{of1}, E_t^{on2}, E_t^{of2}]^\top$. In line with their findings, we choose a training set of $|\mathcal{T}^{in}| = 4320$ (6 months) and a rolling window step of 24 hours.

In addition, we analyze a rolling window model, called LP, that solves exactly (7) and (8) using the full vector of available contextual information. This method is the one utilized in the illustrative example 4.2 but with different inputs. Given the similarities with the other rolling window approach LP2, we also choose a training set length of 6 months and a rolling window step of 24 hours.

Finally, we discuss a benchmark not implementable in practice inspired by the static regret metric defined in (23). We assume perfect information of the whole out-of-sample dataset and leverage problem (7) once more to compute the best linear model in hindsight determined by the vector \mathbf{q}^H . Once this optimal single vector is computed, the whole sequence of offers is determined through $E_t^F = \pi(\mathbf{x}_t, \mathbf{q}^H)$. We name this benchmark FX from FiXed.

Next, we discuss the implementation of OLVN in this case study. The OLVN algorithm does not need to solve an optimization problem but requires initializing two parameters. To choose μ and η , we perform an offline grid search on the chunk of data spanning 07/01/2015 to 12/31/2015. As candidate values for μ we consider $[0, 0.1, \dots, 1]$ and for η we analyze $[10^{-2}, 10^{-3}, 10^{-4}]$. The grid search is carried out executing $3 \times 11 = 33$ independent instances of the OLVN algorithm, initializing each time the OLVN regressor associated with the onshore DK1 forecast to 1 and the rest of the values to 0.01. The average $NV^{os}(\%)$ obtained by each instance is collated in Table 5. After analyzing the results, we select the combination of values $\mu = 0.7$ and $\eta = 0.001$ which achieve the highest $NV^{os}(\%)$. Even though in this case study a grid search was used for the sake of clarity, other more complex cross-validation techniques (Refaeilzadeh et al, 2009) can be used instead to select the values of μ and η , including repeating this process periodically to update the values of μ and η after a change in the environment.

In this case study, we assume a balanced penalty anchor $\bar{\psi}^+ = \bar{\psi}^- = 1$. Again, we initialize the OLVN regressor associated with the onshore DK1 forecast to 1 and the rest of the values to 0.01. In other words, we start the online offering with a strategy very close to FO, mainly relying on the forecast of the wind energy production. We use the next 6 months (01/01/2016 to

06/30/2016) to update (initialize) \mathbf{q}_{OL} with the aim of having a fair comparison against LP and LP2.

The performance of all the methods presented in this section is evaluated using the dataset spanning from 07/01/2016 to 06/04/2021 (5 years with 43,200 samples). The optimization models LP, LP2, and FX are implemented with the Python package Pyomo (Bynum et al, 2021) and solved through the optimization solver CPLEX⁷, whereas the implementation OLNv is developed by the authors based on standard Python packages and uploaded to an open repository⁸.

Table 5: Out-of-sample NV^{OOS} (%) for different combinations of parameters μ and η_0 over the span 07/01/2015 to 12/31/2015. Highlighted in black are shown the best result and parameters selected.

η	0	0.1	0.2	0.3	0.4	μ 0.5	0.6	0.7	0.8	0.9	1
10^{-2}	-13,8	19,2	33,7	19,2	27,7	8,4	39,7	29,2	32,3	32,3	42,0
10^{-3}	12,5	27,1	33,7	36,9	39,2	39,9	42,1	42,2	42,0	41,6	41,5
10^{-4}	-5,2	1,3	4,4	6,0	7,0	7,7	8,2	8,6	8,9	9,4	9,4

5.3 Numerical results

Next, we discuss the results obtained in this case study. We start examining the regret suffered by OLNv over the aforementioned out-of-sample dataset with a length of $D = 43,200$ hours (60 months). Let $\mathcal{T}_j^{\text{OOS}} = \cup_{i=1}^j \mathcal{T}_i^{\text{OOS}}$ and recall $\mathbf{u}_t = \mathbf{q}_i^{\mathcal{H}} \forall t \in \mathcal{T}_i^{\text{OOS}}$. We assess the average dynamic regret R_T^d/T for each sequence $\mathcal{T}_j^{\text{OOS}}, j = 1, \dots, D/l$ with partitions length $l = 2160, 4320, 8640$ hours (3, 6, 12 months). As an additional case, we compute the evolution of the static regret for a sequence $\mathcal{T}_j^{\text{OOS}}, j = 1, \dots, 20$ with a step $l = 2160$ hours (3 months). To this end, in each step we refresh the best single action in hindsight as $\mathbf{q}_j^{\mathcal{H}} = \arg \min_{\mathbf{q} \in Q_j^{\mathcal{H}}} \sum_{t \in \mathcal{T}_j^{\text{OOS}}} NV_t(\mathbf{q})$ and $\mathbf{u}_t = \mathbf{q}_j^{\mathcal{H}} \forall t$.

The four aforementioned regret series are depicted in Figure 5. As expected, the average dynamic regret incurred by OLNv deteriorates quickly as l decreases since lower values of l translate in a more challenging benchmark closer to the the worst-case regret defined in (24). Nevertheless, Figure 5 clearly shows that OLNv achieves a sublinear static regret, i.e., $\lim_{T \rightarrow \infty} \sup \mathcal{R}_T^s/T \leq 0$. This is also the case for the dynamic regret with partitions of length $l \geq 6$ months, proving the ability of OLNv to track dynamic environments.

The economic gains obtained by each method are assessed through the NV^{OOS} (%). The average values achieved over the evaluation dataset are colated in Table 6. First, note that all methods outperform the naive FO strategy

⁷IBM ILOG CPLEX Optimization Studio. See <https://www.ibm.com/analytics/cplex-optimizer>.

⁸Experiment's code and data available at: <https://github.com/Miguel897/online-trading-wind-energy>

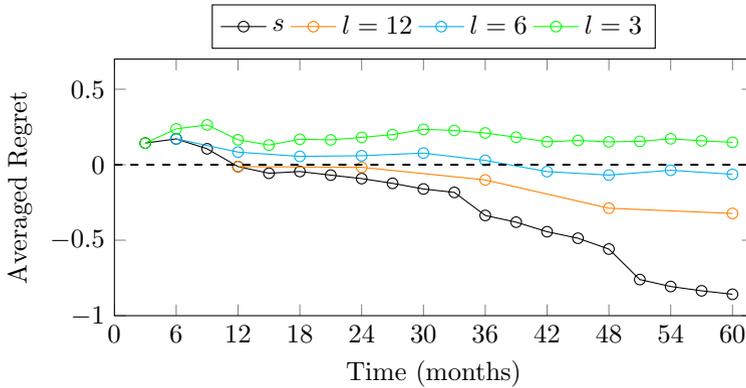


Fig. 5: Average dynamic regret R_T^d/T for $l = 3, 6, 12$, months and static regret R_T^s/T updated every 3 months (denoted as s) of the OLN method.

of offering the DK1 forecast, obtaining positive values and demonstrating that this set of features contributes to reducing the deviation cost.

The LP2 method is developed in a context where recent lags in the penalties are not available. Indeed, the lack of penalty-related features translates into a modest score, showing the evident benefits of disclosing recent information in electricity markets, i.e., reducing the lead time. Even though FX determines the optimal \mathbf{q}^t in hindsight (i.e., under perfect information), its choice is limited to a single vector for the whole horizon. The fact that several approaches perform better than FX proves the dynamic behavior of the uncertain parameters and the need for updating the decision vector. Therefore, it does not come as a surprise that LP improves the first two approaches as it leverages the full vector of features and periodically updates \mathbf{q}_t^{LP} . However, the superior adaptability of OLN allows it to obtain the best score, achieving an additional 7.6% compared to LP and a total 38.6% deviation cost reduction compared to FO. The latter figure translates into an extra 25,930.22 €/year on average for a wind farm with a capacity of 100 MW.

Finally, the last row of Table 6 summarizes the computational time corresponding to four approaches. The FX method requires little time as it only solves a single optimization problem for the whole horizon. This contrasts with the significant amount of time required by the constant re-optimization of LP and LP2. It is noteworthy that even though OLN produces 24 times more updates of the vector \mathbf{q}_t , the time invested is several orders of magnitude lower. In conclusion, OLN is up to the challenge of the electricity markets transformation achieving significant cost reduction together with exceptional computational performance.

Table 6: Out-of-sample NV^{oos} (%) and execution time (s) over the span 07/01/2016 to 06/04/2021.

	LP2	FX	LP	OLNV
NV^{oos} (%)	3.8	24.6	31.0	38.6
Time (s)	23366	53	16077	179

6 Conclusions

This paper develops a new algorithm, named OLNv, combining a variant of the online gradient descent with recent advances that extends the newsvendor model to consider contextual information directly. The component-wise update of the learning rate enables the use of features with different scales seamlessly. In nonstationary environments, conventional stochastic approaches may consider misleading old samples in their training sets. On the contrary, our algorithm tracks the most recent information of the gradients, adapting the learning rate to follow the dynamics of the uncertain parameters and potentially obtaining higher profits. The closed-form expressions derived to compute the projection into the feasible region and a gradient of the objective function yield a efficient algorithm that can be used in computationally expensive problems. We envision the use of OLNv in future electricity markets that evolves toward continuous offering with reduced lead time. In particular, we apply this algorithm to the wind farm problem offering in an hourly forward market with a dual-price settlement for imbalances.

Several numerical experiments are carried out to assess the properties of the proposed OLNv algorithm. In the first illustrative example, we compare the behavior of two alternative implementations, namely, a subgradient approach and a smooth approximation of the original newsvendor function. The numerical and theoretical analysis provided in this example indicates that computing subgradient on the original objective function proves more profitable since it avoids update errors that may be introduced by the smooth approximation. Consequently, we determined that the subgradient implementation was the most suitable to this application and used it throughout the rest of the numerical experiments. Nevertheless, the smooth approximation could be utilized in other applications where other technical concerns advice a smooth update.

The second example shows the adaptability of the OLNv algorithm to nonstationary environments, clearly outperforming other stochastic approaches that optimize (using mathematical programming techniques) over a training set of past information. This superior performance is justified by the point-wise update that only uses the most recent information. Our case study, built upon real data of the Danish TSO Energinet, shows that OLNv achieves a 38.6% cost reduction against using a point forecast as offer and 7.6% compared to a state-of-the-art method. These significant improvements contribute to accelerating the integration of renewable energy technologies. Furthermore, we empirically analyze several dynamic definitions of regret, showing the desired sublinear convergence against most benchmarks.

Although this research focused on wind energy producers, OLVN is readily applicable to managing a portfolio of variable renewable energies with zero marginal cost, including wind, solar and other technologies. Similar algorithms can be developed when the producer's portfolio includes other assets such as loads, thermal power plants, or energy storage facilities, replacing the aggregated source of uncertainty, i.e., the variable net energy production, by a linear decision rule. In this case, the projection step on the feasible region would likely involve solving a quadratic optimization program that can still be efficiently solved with modern solvers, when the feasible region is convex. Another attractive front is extending the OLVN algorithm to address inter-temporal constraints, observing a similar note with regard to the feasible region as in the previous case. This may require first generalizing the newsvendor framework to offering in electricity markets though.

Future work also includes delving into the theoretical guarantees that this algorithm offers in terms of regret. On a different front, a wealth of other algorithms within the field of online learning can be applied to this problem, potentially bringing additional benefits such as faster convergence rates or improved performance. Similarly, variable selection techniques could help determine the subset of the available feature streams that provide the most economic value, whereas nonlinear mapping, i.e., kernels or generalized additive models (GAMs), can extend the regression capabilities of the method. Another exciting line of research concerns the risk analysis of the producer, where other metrics can be used instead of the expected value to create risk-averse strategies.

Acknowledgments

M. Á. Muñoz is funded by the Spanish Ministry of Science, Innovation and Universities through the State Training Subprogram 2018 of the State Program for the Promotion of Talent and its Employability in R&D&I, within the framework of the State Plan for Scientific and Technical Research and Innovation 2017-2020 and by the European Social Fund.

P. Pinson and J. Kazempour are partly supported through the Smart4RES project (European Union's Horizon 2020, No. 864337). The sole responsibility of this publication lies with the authors. The European Union is not responsible for any use that may be made of the information contained therein.

Appendix A Smooth function properties

This appendix provides a lemma and several propositions related to the smooth approximation $NV_{t,\alpha}$ defined in eq. (10). Some of the proofs in this appendix are based on the proofs provided in Zheng (2011). In this appendix we assume that $\psi_t^+, \psi_t^- \geq 0$ and $\psi_t^+ + \psi_t^- > 0 \forall t$. Next, we define an auxiliary function

$S_{t,\alpha}(u)$, $S_{t,\alpha} : \mathbb{R} \rightarrow \mathbb{R}$ with $\alpha > 0$ as follows

$$S_{t,\alpha}(u) = \psi_t^+ u + \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{-u/\alpha}), \quad (\text{A1})$$

where $u \in \mathbb{R}$. We leverage this function in the proofs of this appendix. First, we prove the convexity of $S_{t,\alpha}$.

Lemma 1. *For any given $\alpha > 0$, the function $S_{t,\alpha}$, defined in (A1), is a convex function.*

Proof From the definition of $S_{t,\alpha}$ in eq. (A1), we calculate that

$$\frac{d^2 S_{t,\alpha}(u)}{du^2} = \frac{\psi_t^+ + \psi_t^-}{\alpha} \frac{e^{-\frac{u}{\alpha}}}{(1 + e^{-\frac{u}{\alpha}})^2} > 0 \quad (\text{A2})$$

for any $u \in \mathbb{R}$ since $\psi_t^+ + \psi_t^- > 0$ and $\alpha > 0$.

We use this intermediate result to prove the convexity of $NV_{t,\alpha}$ in the following Proposition.

Proposition 1. *For any given $\alpha > 0$, the function $NV_{t,\alpha}$, defined in (10), is a convex function of \mathbf{q} .*

Proof Let $u = E_t - \mathbf{x}_t^\top \mathbf{q}$ in (A1). Thus,

$$NV_{t,\alpha}(\mathbf{q}) = S_{t,\alpha}(E_t - \mathbf{x}_t^\top \mathbf{q}). \quad (\text{A3})$$

For $0 \leq \omega \leq 1$ and any \mathbf{q}_1 and \mathbf{q}_2 , we have

$$\begin{aligned} NV_{t,\alpha}(\omega \mathbf{q}_1 + (1 - \omega) \mathbf{q}_2) &= S_{t,\alpha}(E_t - \mathbf{x}_t^\top (\omega \mathbf{q}_1 + (1 - \omega) \mathbf{q}_2)) \\ &= S_{t,\alpha}(E_t - \omega \mathbf{x}_t^\top \mathbf{q}_1 - (1 - \omega) \mathbf{x}_t^\top \mathbf{q}_2) \\ &= S_{t,\alpha}(\omega(E_t - \mathbf{x}_t^\top \mathbf{q}_1) - (1 - \omega)(E_t - \mathbf{x}_t^\top \mathbf{q}_2)) \quad (\text{A4}) \\ &\leq \omega S_{t,\alpha}(E_t - \mathbf{x}_t^\top \mathbf{q}_1) + (1 - \omega) S_{t,\alpha}(E_t - \mathbf{x}_t^\top \mathbf{q}_2), \quad (\text{A5}) \end{aligned}$$

where the inequality in eq. (A5) follows from the convexity of $S_{t,\alpha}$, proved in Lemma 1. Then, leveraging eq. (A3), the above inequality renders

$$NV_{t,\alpha}(\omega \mathbf{q}_1 + (1 - \omega) \mathbf{q}_2) \leq \omega NV_{t,\alpha}(\mathbf{q}_1) + (1 - \omega) NV_{t,\alpha}(\mathbf{q}_2), \quad (\text{A6})$$

proving that $NV_{t,\alpha}$ is a convex function on \mathbf{q} .

Next, we show that $NV_{t,\alpha}$ asymptotically approaches NV_t for $\alpha \rightarrow 0$. We also show that the function $NV_{t,\alpha}$ upper bounds NV_t for all $q \in \mathbb{R}^p$.

Proposition 2. *Let NV_t and $NV_{t,\alpha}$ be the functions defined in eq. (9) and (10), in that order, with $\alpha > 0$. Then, we have*

$$0 < NV_{t,\alpha}(\mathbf{q}) - NV_t(\mathbf{q}) \leq \alpha(\psi_t^+ + \psi_t^-) \log 2 \quad (\text{A7})$$

for any $\mathbf{q} \in \mathbb{R}^p$. Thus,

$$\lim_{\alpha \rightarrow 0^+} NV_{t,\alpha}(\mathbf{q}) = NV_t(\mathbf{q}). \quad (\text{A8})$$

Proof When $E_t - \mathbf{x}_t^\top \mathbf{q} \geq 0$, we have that

$$NV_{t,\alpha}(\mathbf{q}) - NV_t(\mathbf{q}) = \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{-(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha}), \quad (\text{A9})$$

hence,

$$0 < NV_{t,\alpha}(\mathbf{q}) - NV_t(\mathbf{q}) \leq \alpha(\psi_t^+ + \psi_t^-) \log 2 \quad (\text{A10})$$

for $E_t - \mathbf{x}_t^\top \mathbf{q} \geq 0$. When $E_t - \mathbf{x}_t^\top \mathbf{q} < 0$,

$$\begin{aligned} NV_{t,\alpha}(\mathbf{q}) - NV_t(\mathbf{q}) &= (\psi_t^+ + \psi_t^-)(E_t - \mathbf{x}_t^\top \mathbf{q}) \\ &\quad + \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{-(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha}) \end{aligned} \quad (\text{A11})$$

$$= \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha}). \quad (\text{A12})$$

While

$$\begin{aligned} 0 &< \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{(E_t - \mathbf{x}_t^\top \mathbf{q})/\alpha}) \\ &< \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{0/\alpha}) = \alpha(\psi_t^+ + \psi_t^-) \log 2, \end{aligned} \quad (\text{A13})$$

since $E_t - \mathbf{x}_t^\top \mathbf{q} < 0$. This shows that $NV_{t,\alpha}(\mathbf{q}) - NV_t(\mathbf{q})$ also falls in the range $(0, \alpha(\psi_t^+ + \psi_t^-) \log 2)$ for $E_t - \mathbf{x}_t^\top \mathbf{q} < 0$. Thus, eq. (A7) is proved. Eq. (A8) follows directly by letting $\alpha \rightarrow 0^+$ in eq. (A7).

Finally, we show that for high values of $|E_t - \mathbf{x}_t^\top \mathbf{q}|$ the function $NV_{t,\alpha}$ asymptotically approximate NV_t .

Proposition 3. *Let NV_t and $NV_{t,\alpha}$ be the functions defined in eq. (9) and (10), in that order, with $\alpha > 0$. Then, when $|E_t - \mathbf{x}_t^\top \mathbf{q}| \rightarrow \infty$ we have that $NV_{t,\alpha} - NV_t \rightarrow 0$.*

Proof For the sake of a clearer exposition we define $\mu(\mathbf{q}) = E_t - \mathbf{x}_t^\top \mathbf{q}$, where $\mu: \mathbb{R}^p \rightarrow \mathbb{R}$. When $\mu(\mathbf{q}) \rightarrow +\infty$, we have that

$$\lim_{\mu(\mathbf{q}) \rightarrow +\infty} NV_{t,\alpha}(\mathbf{q}) - NV_t(\mathbf{q}) = \lim_{\mu(\mathbf{q}) \rightarrow +\infty} \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{-(\mu(\mathbf{q}))/\alpha}) = 0 \quad (\text{A14})$$

When $\mu(\mathbf{q}) \rightarrow -\infty$, leveraging eq. (A12) we have that

$$\lim_{\mu(\mathbf{q}) \rightarrow -\infty} NV_{t,\alpha}(\mathbf{q}) - NV_t(\mathbf{q}) = \lim_{\mu(\mathbf{q}) \rightarrow -\infty} \alpha(\psi_t^+ + \psi_t^-) \log(1 + e^{(\mu(\mathbf{q}))/\alpha}) = 0 \quad (\text{A15})$$

Combining both cases, this proposition is proved.

References

- Ban GY, Rudin C (2019) The big data newsvendor: Practical insights from machine learning. *Operations Research* 67(1):90–108
- Bashir AA, Lehtonen M (2018) Day-ahead rolling window optimization of islanded microgrid with uncertainty. In: 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), IEEE, pp 1–6
- Bertsimas D, Kallus N (2019) From predictive to prescriptive analytics. *Management Science* 66(3):1025–1044
- Besbes O, Gur Y, Zeevi A (2015) Non-stationary stochastic optimization. *Operations Research* 63(5):1227–1244
- Bremnes JB (2004) Probabilistic wind power forecasts using local quantile regression. *Wind Energy* 7(1):47–54
- Bynum ML, Hackebeil GA, Hart WE, et al (2021) *Pyomo—optimization modeling in Python*, vol 67, 3rd edn. Springer Science & Business Media
- Colombino M, Dall’Anese E, Bernstein A (2019) Online optimization as a feedback controller: Stability and tracking. *IEEE Transactions on Control of Network Systems* 7(1):422–432
- Conejo A, Carrión M, Morales J (2010) *Decision-making under Uncertainty in Electricity Markets*. Springer, New York, NY
- Dent C, Bialek J, Hobbs B (2011) Opportunity cost bidding by wind generators in forward markets: Analytical results. *IEEE Transactions on Power Systems* 26(3):1600–1608
- Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12(61):2121–2159
- Gan L, Low SH (2016) An online gradient algorithm for optimal power flow on radial networks. *IEEE Journal on Selected Areas in Communications* 34(3):625–638

- Guo Z, Pinson P, Chen S, et al (2021) Online optimization for real-time peer-to-peer electricity market mechanisms. *IEEE Transactions on Smart Grid* 12(5):4151–4163
- Hargreaves JJ, Hobbs BF (2012) Commitment and dispatch with uncertain wind generation by dynamic programming. *IEEE Transactions on Sustainable Energy* 3(4):724–734
- Hauswirth A, Bolognani S, Hug G, et al (2016) Projected gradient descent on riemannian manifolds with applications to online power system optimization. In: 2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, pp 225–232
- Hauswirth A, Zanardi A, Bolognani S, et al (2017) Online optimization in closed loop on the power flow manifold. In: 2017 IEEE Manchester PowerTech, IEEE, pp 1–6
- Hazan E, et al (2016) Introduction to online convex optimization. *Foundations and Trends® in Optimization* 2(3-4):157–325
- Kuznetsova E, Li YF, Ruiz C, et al (2013) Reinforcement learning for microgrid energy management. *Energy* 59:133–146
- Liakopoulos N, Destounis A, Paschos G, et al (2019) Cautious regret minimization: Online optimization with long-term budget constraints. In: International Conference on Machine Learning, PMLR, pp 3944–3952
- Maggioni F, Cagnolari M, Bertazzi L (2019) The value of the right distribution in stochastic programming with application to a newsvendor problem. *Computational Management Science* 16(4):739–758
- Mazzi N, Pinson P (2016) Purely data-driven approaches to trading of renewable energy generation. In: 13th International Conference on the European Energy Market (EEM), pp 1–5
- McMahan HB, Streeter M (2010) Adaptive bound optimization for online convex optimization. *arXiv preprint arXiv:10024908*
- Morales J, Conejo A, Pérez-Ruiz J (2010) Short-term trading for a wind power producer. *IEEE Transactions on Power Systems* 25(1):554–564
- Morales J, Conejo A, Madsen H, et al (2014) *Integrating Renewables in Electricity Markets – Operational Problems*. Springer, New York, NY
- Muñoz M, Morales JM, Pineda S (2020) Feature-driven improvement of renewable energy forecasting and trading. *IEEE Transactions on Power Systems* 35(5):3753–3763

- Narayanaswamy B, Garg VK, Jayram T (2012) Online optimization for the smart (micro) grid. In: Proceedings of the 3rd International Conference on Future Energy Systems, pp 1–10
- Nonhoff M, Müller MA (2020) Online gradient descent for linear dynamical systems. *IFAC-PapersOnLine* 53(2):945–952
- Orabona F (2022) A modern introduction to online learning. arXiv preprint arXiv:1912.13213
- Pinson P, Chevallier C, Kariniotakis GN (2007) Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems* 22(3):1148–1156
- Qin Y, Wang R, Vakharia AJ, et al (2011) The newsvendor problem: Review and directions for future research. *European Journal of Operational Research* 213(2):361–374
- Refaeilzadeh P, Tang L, Liu H (2009) Cross-validation. *Encyclopedia of Database Systems* 5:532–538
- Shalev-Shwartz S, et al (2012) Online learning and online convex optimization. *Foundations and Trends[®] in Machine Learning* 4(2):107–194
- Sherman U, Koren T (2021) Lazy OCO: Online convex optimization on a switching budget. In: Conference on Learning Theory, PMLR, pp 3972–3988
- Stratigakos A, Camal S, Michiorri A, et al (2022) Prescriptive trees for integrated forecasting and optimization applied in trading of renewable energy. *IEEE Transactions on Power Systems*, 37(6):4696–4708
- Sutton RS, Barto AG (2018) Reinforcement learning: An introduction. MIT Press
- Van Parys BP, Esfahani PM, Kuhn D (2021) From data to decisions: Distributionally robust optimization is optimal. *Management Science* 67(6):3387–3402
- Wood K, Bianchin G, Dall’Anese E (2021) Online projected gradient descent for stochastic optimization with decision-dependent distributions. *IEEE Control Systems Letters* 6:1646–1651
- Yuan D, Bhardwaj A, Petersen I, et al (2022) Towards online optimization for power grids. *ACM SIGEnergy Energy Informatics Review* 1(1):51–58
- Zeiler MD (2012) ADADELTA: An adaptive learning rate method. arXiv preprint arXiv:1212.5701

- Zheng S (2011) Gradient descent algorithms for quantile regression with smooth approximation. *International Journal of Machine Learning and Cybernetics* 2(3):191–207
- Zinkevich M (2003) Online convex programming and generalized infinitesimal gradient ascent. In: *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp 928–936
- Zugno M, Jónsson T, Pinson P (2013) Trading wind energy on the basis of probabilistic forecasts both of wind generation and of market quantities. *Wind Energy* 16(6):909–926